Available online at www.sciencedirect.com
ScienceDirect
Stochastic Processes and their Applications 117 (2007) 1316–1329
stochastic processes and their applications
www.elsevier.com/locate/spa

Estimation for the additive Gaussian channel and Monge–Kantorovitch measure transportation

Ali Süleyman Üstünel*

ENST, Infres, 46 Rue Barrault, 75013 Paris, France

Received 26 September 2006; received in revised form 17 January 2007; accepted 20 January 2007
Available online 25 January 2007

Abstract

Let (W, μ, H) be an abstract Wiener space and assume that Y is a signal of the form $Y = X + w$, where X is an H-valued random variable, w is the generic element of W. Under the hypothesis of independence of w and X, we show that the quadratic estimate of X, denoted by $\hat{X}(Y) = E[X|Y]$, is of the form $\nabla F(Y)$, where F is an H-convex function on W. We prove also some relations between the quadratic estimate error and the Wasserstein distance between some natural probabilities induced by the shift $I_H + \nabla F$ and the conditional law of Y given X.

MSC: 60H07; 60H05; 60G35; 46G12; 47H05; 35J60

Keywords: Gaussian channel; Malliavin calculus; Monge–Kantorovitch (Kantorovich) measure transportation; H-convex functionals; Wasserstein distance

1. Introduction

This paper is devoted to the problem of estimation of an additive Gaussian channel in the framework of an infinite dimensional Fréchet space as proposed in a recent work of Zakai [14] (although in [14], the general setting is a Banach space, everything goes through for a separable Fréchet space) and its relations to the Monge–Kantorovitch measure transportation problem. Let W be a separable Fréchet space. We assume that W supports a Gaussian measure μ with a Cameron–Martin space, denoted by H. Assume that X is an H-valued random variable defined

* Tel.: +33 145817267; fax: +33 145813119.
E-mail address: ustunel@enst.fr.

0304-4149/$ - see front matter
doi:10.1016/j.spa.2007.01.007

on an arbitrary probability space rich enough in such a way that one has also a W-valued random variable J whose law is equal to μ. Under the assumption of the independence of X and J, one wants to calculate the quadratic estimate of X with respect to the observation Y defined as

$$Y = X + J.$$

Assuming

$$E[\exp \varepsilon |X|_H] < \infty,$$

for some $\varepsilon > 0$, we prove that the Radon–Nikodym density of the law of Y, denoted by μ_Y, is of the form $\exp F$, where F is an H-analytic Wiener functional on W, besides which it is H-convex (cf. [3]) and the quadratic estimate of X w.r.t. Y, i.e., $E[X|Y]$ can be expressed as $\nabla F(Y)$, where ∇ denotes the Sobolev derivative on the abstract Wiener space (W, μ, H). Let us recall that the solutions of the Monge–Kantorovitch problem on the Wiener space (with respect to the singular Cameron–Martin distance) are the images of the initial measures under the maps of the form $(I_W \times \nabla K)$, where K is a 1-convex Wiener functional (cf. [4–6]). Since H-convexity implies 1-convexity (cf. [3]), it is clear that the estimation problem is related to the Monge–Kantorovitch problem for some measures. In this paper we expose these measures and also give some further estimation results which correspond to the situation in which the signal process originates from a space on which a stochastic analysis in the sense of the Malliavin calculus exists.

2. Preliminaries and notation

Let W be a separable Fréchet space equipped with a Gaussian measure μ of zero mean whose support is the whole space. The corresponding Cameron–Martin space is denoted by H. Recall that the injection $H \hookrightarrow W$ is compact and its adjoint is the natural injection $W^\star \hookrightarrow H^\star \subset L^2(\mu)$. The triple (W, μ, H) is called an abstract Wiener space. Recall that $W = H$ if and only if W is finite dimensional. A subspace F of H is called regular if the corresponding orthogonal projection has a continuous extension to W, denoted again by the same letter. It is well known that there exists an increasing sequence of regular subspaces $(F_n, n \geq 1)$, called total, such that $\cup_n F_n$ is dense in H and in W. Let $\sigma(\pi_{F_n})$[1] be the σ-algebra generated by π_{F_n}; then for any $f \in L^p(\mu)$, the martingale sequence $(E[f|\sigma(\pi_{F_n})], n \geq 1)$ converges to f (strongly if $p < \infty$) in $L^p(\mu)$. Observe that the function $f_n = E[f|\sigma(\pi_{F_n})]$ can be identified with a function on the finite dimensional abstract Wiener space (F_n, μ_n, F_n), where $\mu_n = \pi_n \mu$.

Since the translations of μ with the elements of H induce measures equivalent to μ, the Gâteaux derivative in the H direction of the random variables is a closable operator on $L^p(\mu)$-spaces and this closure will be denoted by ∇; cf., for example, [2,11,12]. The corresponding Sobolev spaces (the equivalence classes) of the real random variables will be denoted as $\mathbb{D}_{p,k}$, where $k \in \mathbb{N}$ is the order of differentiability and $p > 1$ is the order of integrability. If the random variables have values in some separable Hilbert space, say Φ, then we shall define similarly the corresponding Sobolev spaces and they are denoted as $\mathbb{D}_{p,k}(\Phi)$, $p > 1, k \in \mathbb{N}$. Since $\nabla : \mathbb{D}_{p,k} \to \mathbb{D}_{p,k-1}(H)$ is a continuous and linear operator its adjoint is a well-defined operator which we represent by δ. In the case of classical Wiener space, i.e., when $W = C(\mathbb{R}_+, \mathbb{R}^d)$, then δ coincides with the Itô integral of the Lebesgue density of the adapted elements of $\mathbb{D}_{p,k}(H)$ (cf. [11,12]).

[1] For notational simplicity, in the sequel we shall denote it by π_n.

For any $t \geq 0$ and measurable $f : W \to \mathbb{R}_+$, we define

$$P_t f(x) = \int_W f(e^{-t}x + \sqrt{1 - e^{-2t}}\, y)\mu(dy),$$

and it is well known that $(P_t, t \in \mathbb{R}_+)$ is a hypercontractive semigroup on $L^p(\mu)$, $p > 1$, which is called the Ornstein–Uhlenbeck semigroup (cf. [2,8,11,12]). Its infinitesimal generator is denoted by $-\mathcal{L}$ and we call $\mathcal{L}$ the Ornstein–Uhlenbeck operator (sometimes called the number operator by physicists). The norms defined by

$$\|\phi\|_{p,k} = \|(I + \mathcal{L})^{k/2}\phi\|_{L^p(\mu)} \tag{2.1}$$

are equivalent to the norms defined by the iterates of the Sobolev derivative ∇. This observation permits us to identify the duals of the space $\mathbb{D}_{p,k}(\Phi)$; $p > 1, k \in \mathbb{N}$ by $\mathbb{D}_{q,-k}(\Phi')$, with $q^{-1} = 1 - p^{-1}$, where the latter space is defined by replacing k in (2.1) by $-k$; this gives us the distribution spaces on the Wiener space W (in fact we can take as k any real number). An easy calculation shows that, formally, $\delta \circ \nabla = \mathcal{L}$, and this permits us to extend the divergence and the derivative operators to the distributions as linear, continuous operators. In fact $\delta : \mathbb{D}_{q,k}(H \otimes \Phi) \to \mathbb{D}_{q,k-1}(\Phi)$ and $\nabla : \mathbb{D}_{q,k}(\Phi) \to \mathbb{D}_{q,k-1}(H \otimes \Phi)$ continuously, for any $q > 1$ and $k \in \mathbb{R}$, where $H \otimes \Phi$ denotes the completed Hilbert–Schmidt tensor product (cf., for instance, [11,12]).

Let us recall some facts from convex analysis. Let K be a Hilbert space; a subset S of $K \times K$ is called cyclically monotone if any finite subset $\{(x_1, y_1), \ldots, (x_N, y_N)\}$ of S satisfies the following algebraic condition:

$$\langle y_1, x_2 - x_1\rangle + \langle y_2, x_3 - x_2\rangle + \cdots + \langle y_{N-1}, x_N - x_{N-1}\rangle + \langle y_N, x_1 - x_N\rangle \leq 0,$$

where $\langle \cdot, \cdot \rangle$ denotes the inner product of K. It turns out that S is cyclically monotone if and only if

$$\sum_{i=1}^N (y_i, x_{\sigma(i)} - x_i) \leq 0,$$

for any permutation σ of $\{1, \ldots, N\}$ and for any finite subset $\{(x_i, y_i) : i = 1, \ldots, N\}$ of S. Note that S is cyclically monotone if and only if any translate of it is cyclically monotone. By a theorem of Rockafellar, any cyclically monotone set is contained in the graph of the subdifferential of a convex function in the sense of convex analysis [9] and even if the function may not be unique its subdifferential is unique.

Let now (W, μ, H) be an abstract Wiener space; a measurable function $f : W \to \mathbb{R} \cup \{\infty\}$ is called 1-convex if the map

$$h \to f(x + h) + \frac{1}{2}|h|_H^2 = F(x, h)$$

is convex on the Cameron–Martin space H with values in $L^0(\mu)$. Note that this notion is compatible with the μ-equivalence classes of random variables thanks to the Cameron–Martin theorem. It is proven in [4] that this definition is equivalent to the following condition: Let $(\pi_n, n \geq 1)$ be a sequence of regular, finite dimensional, orthogonal projections of H, increasing to the identity map I_H. Denote also by π_n its continuous extension to W and define

$\pi_n^{\perp} = I_W - \pi_n$. For $x \in W$, let $x_n = \pi_n x$ and $x_n^{\perp} = \pi_n^{\perp} x$. Then f is 1-convex if and only if

$$x_n \to \frac{1}{2}|x_n|_H^2 + f(x_n + x_n^{\perp})$$

is $\pi_n^{\perp}\mu$-almost surely convex.

2.1. Preliminaries concerning the Monge–Kantorovitch measure transportation problem

Definition 1. Let ξ and η be two probabilities on $(W, \mathcal{B}(W))$. We say that a probability γ on $(W \times W, \mathcal{B}(W \times W))$ is a solution of the Monge–Kantorovitch problem associated with the couple (ξ, η) if the first marginal of γ is ξ, the second one is η and if

$$J(\gamma) = \int_{W\times W} |x-y|_H^2 \mathrm{d}\gamma(x,y) = \inf\left\{\int_{W\times W} |x-y|_H^2 \mathrm{d}\beta(x,y) : \beta \in \Sigma(\xi,\eta)\right\},$$

where $\Sigma(\xi, \eta)$ denotes the set of all the probability measures on $W \times W$ whose first and second marginals are respectively ξ and η. We shall denote the Wasserstein distance between ξ and η, which is the positive square root of this infimum, as $d_H(\xi, \eta)$.

Remark: By the weak compactness of probability measures on $W \times W$ and the lower semi-continuity of the strictly convex cost function, the infimum in the definition is attained even if the functional J is identically infinity. In this latter case we say that the solution is degenerate.

The next result, which is an extension (cf. [4–6]) of the finite dimensional version due to Talagrand [10], gives a sufficient condition for the finiteness of the Wasserstein distance in the case where one of the measures is the Wiener measure μ and the second one is absolutely continuous with respect to it. We give a short proof for the sake of completeness:

Theorem 1. *Let $L \in \mathbb{L}\log\mathbb{L}(\mu)$ be a positive random variable with $E[L] = 1$*[2] *and let ν be the measure $\mathrm{d}\nu = L\mathrm{d}\mu$. We then have*

$$\mathrm{d}_H^2(\nu,\mu) \le 2E[L\log L]. \tag{2.2}$$

Proof. Let us remark first that we can take W as the classical Wiener space $W = C_0([0, 1])$ and, using the stopping techniques of the martingale theory, we may assume that L is upper and lower bounded almost surely. Then a classical result of the Itô calculus implies that L can be represented as an exponential martingale

$$L_t = \exp\left\{-\int_0^t \dot{u}_\tau \mathrm{d}W_\tau - \frac{1}{2}\int_0^t |\dot{u}_\tau|^2 \mathrm{d}\tau\right\},$$

with $L = L_1$, where $(\dot{u}_t, t \in [0, 1])$ is a measurable process adapted to the filtration of the canonical Wiener process $(t, x) \to W_t(x) = x(t)$. Let us define $u : W \to H$ as $u(t, x) = \int_0^t \dot{u}_\tau(x)\mathrm{d}\tau$ and $U : W \to W$ as $U(x) = x + u(x)$. The Girsanov theorem implies that $x \to U(x)$ is a Brownian motion under ν; hence the image of the measure ν under the map $U \times I_W : W \to W \times W$ denoted by $\beta = (U \times I)\nu$ belongs to $\Sigma(\mu, \nu)$. Let γ be any optimal

[2] In the sequel we denote the expectation w.r.t. the Wiener measure by E.

measure; then

$$\begin{aligned} J(\gamma) = \mathrm{d}_H^2(\nu, \mu) &\leq \int_{W\times W} |x-y|_H^2 \mathrm{d}\beta(x, y) \\ &= E[|u|_H^2 L] \\ &= 2E[L \log L], \end{aligned}$$

where the last equality follows also from the Girsanov theorem and the Itô stochastic calculus. □

The next two theorems which explain the existence and several properties of the solutions of the Monge–Kantorovitch problem and the transport maps have been proven in [1,5–7].

Theorem 2 (*General Case*). *Suppose that ρ and ν are two probability measures on W such that*

$$d_H(\rho, \nu) < \infty.$$

Let $(\pi_n, n \geq 1)$ be a total increasing sequence of regular projections (of H, converging to the identity map of H). Suppose that, for any $n \geq 1$, the regular conditional probabilities $\rho(\cdot|\pi_n^\perp = x^\perp)$ vanish $\pi_n^\perp\rho$-almost surely on the subsets of $(\pi_n^\perp)^{-1}(W)$ with Hausdorff dimension $n-1$. Then there exists a unique solution of the Monge–Kantorovitch problem, denoted by $\gamma \in \Sigma(\rho, \nu)$ and γ is supported by the graph of a Borel map T which is the solution of the Monge problem. $T : W \to W$ is of the form $T = I_W + \xi$, where $\xi \in H$ almost surely. Besides we have

$$\begin{aligned} \mathrm{d}_H^2(\rho, \nu) &= \int_{W\times W} |T(x) - x|_H^2 \mathrm{d}\gamma(x, y) \\ &= \int_W |T(x) - x|_H^2 \mathrm{d}\rho(x), \end{aligned}$$

and for $\pi_n^\perp\rho$-almost almost all $x_n^\perp$, the map $u \to \xi(u + x_n^\perp)$ is cyclically monotone on $(\pi_n^\perp)^{-1}\{x_n^\perp\}$, in the sense that

$$\sum_{i=1}^{N} \left(\xi(x_n^\perp + u_i), u_{i+1} - u_i\right)_H \leq 0$$

$\pi_n^\perp\rho$-almost surely, for any cyclic sequence $\{u_1, \ldots, u_N, u_{N+1} = u_1\}$ from $\pi_n(W)$. Finally, if, for any $n \geq 1$, $\pi_n^\perp\nu$-almost surely, $\nu(\cdot|\pi_n^\perp = y^\perp)$ also vanishes on the $n-1$-Hausdorff dimensional subsets of $(\pi_n^\perp)^{-1}(W)$, then T is invertible, i.e., there exists $S : W \to W$ of the form $S = I_W + \eta$ such that $\eta \in H$ satisfies a similar cyclic monotonicity property to ξ and that

$$\begin{aligned} 1 &= \gamma\left\{(x, y) \in W \times W : T \circ S(y) = y\right\} \\ &= \gamma\left\{(x, y) \in W \times W : S \circ T(x) = x\right\}. \end{aligned}$$

In particular we have

$$\begin{aligned} \mathrm{d}_H^2(\rho, \nu) &= \int_{W\times W} |S(y) - y|_H^2 \mathrm{d}\gamma(x, y) \\ &= \int_W |S(y) - y|_H^2 \mathrm{d}\nu(y). \end{aligned}$$

Remark 1. In particular, for all the measures ρ which are absolutely continuous with respect to the Wiener measure μ, the second hypothesis is satisfied, i.e., the measure $\rho(\cdot|\pi_n^\perp = x_n^\perp)$ vanishes on the sets of Hausdorff dimension $n-1$.

The case where one of the measures is the Wiener measure and the other is absolutely continuous with respect to μ is the most important one for the applications. Consequently we give the related results separately in the following theorem where the tools of the Malliavin calculus give more information about the maps ξ and η of Theorem 2:

Theorem 3 (*Gaussian Case*). *Let ν be the measure $\mathrm{d}\nu = L\mathrm{d}\mu$, where L is a positive random variable, with $E[L] = 1$. Assume that $d_H(\mu, \nu) < \infty$ (for instance $L \in \mathbb{L}\log\mathbb{L}$). Then there exists a 1-convex function $\phi \in \mathbb{D}_{2,1}$, unique up to a constant, such that the map $T = I_W + \nabla\phi$ is the unique solution of the original problem of Monge. Moreover, its graph supports the unique solution of the Monge–Kantorovitch problem γ. Consequently*

$$(I_W \times T)\mu = \gamma.$$

In particular T maps μ to ν and T is almost surely invertible, i.e., there exists some T^{-1} such that $T^{-1}\nu = \mu$ and that

$$\begin{aligned} 1 &= \mu\{x : T^{-1} \circ T(x) = x\} \\ &= \nu\{y \in W : T \circ T^{-1}(y) = y\}. \end{aligned}$$

Remark 2. Assume that the operator ∇ is closable with respect to ν; then we have $\eta = \nabla\psi$. In particular, if ν and μ are equivalent, then we have

$$T^{-1} = I_W + \nabla\psi,$$

where is ψ is a 1-convex function. ψ is called the dual potential of the MKP (μ, ν) and we have the following relations:

$$\phi(x) + \psi(y) + \frac{1}{2}|x - y|_H^2 \geq 0,$$

for any $x, y \in W$, and

$$\phi(x) + \psi(y) + \frac{1}{2}|x - y|_H^2 = 0$$

γ-almost surely.

Remark 3. Let $(e_n, n \in \mathbb{N})$ be complete, orthonormal in H; denote by V_n the sigma algebra generated by $\{\delta e_1, \ldots, \delta e_n\}$ and let $L_n = E[L|V_n]$. If $\phi_n \in \mathbb{D}_{2,1}$ is the function constructed in Theorem 3, corresponding to L_n, then, using the inequality (2.2), we can prove that the sequence $(\phi_n, n \in \mathbb{N})$ converges to ϕ in $\mathbb{D}_{2,1}$.

3. Calculations of estimates and their regularity, and relations with the Monge–Kantorovitch measure transportation

Assume that X is an H-valued random variable defined on some probability space $(\Omega', \mathcal{A}', P')$; take $\Omega = \Omega' \times W$, $P = P' \times \mu$ and $J(\theta, w) = w$ so that X is independent of J. In the sequel, for typographical reasons, we shall denote the r.v. $J(\theta, w)$ with the same notation as the generic element of W, i.e., as w, since $J(\theta, \cdot)$ is equal to the identity I_W of W; this should not create any ambiguity. Y denotes the observation defined as

$$Y = X + w.$$

Note that the law of Y, denoted by μ_Y, is absolutely continuous with respect to μ (in fact they are equivalent) and the corresponding Radon–Nikodym derivative is given as

$$\begin{aligned} l(w) &= \frac{\mathrm{d}\mu_Y}{\mathrm{d}\mu}(w) \\ &= \int_H \exp\left\{(x, w)_H - \frac{1}{2}|x|_H^2\right\} \alpha(\mathrm{d}x), \end{aligned}$$

where $(x, w)_H$ is defined as the limit in probability of the random variables $((\pi_n w, x)_H, n \geq 1)$, where the probability is the product measure $\alpha \times \mu$, α being the law of X, and $(x, w)_H$ is defined as the limit in probability.

In the sequel we shall assume that X is exponentially integrable, i.e., the existence of an $\varepsilon > 0$ for which

$$\int_H \exp[\varepsilon |x|_H] \mathrm{d}\alpha(x) < \infty,$$

holds true.

Proposition 1. *The density l has a modification which is almost surely real H-analytic.*

Proof. Let us denote by $L(x, w)$ the conditional density

$$L(x, w) = \exp\left[(x, w)_H - \frac{1}{2}|x|_H^2\right].$$

We have

$$l(w) = \int_H L(x, w)\alpha(\mathrm{d}x),$$

and therefore

$$\begin{aligned} E[|\nabla^k l|_{H^{\otimes k}}] &\leq E \int_H |x|_H^k L(x, w) \mathrm{d}\alpha(x) \\ &= \int_H |x|_H^k \mathrm{d}\alpha(x). \end{aligned}$$

Hence

$$\sum_{k=0}^{\infty} \frac{1}{k!} \|\nabla^k l\|_{L^1(\mu, H^{\otimes k})} < \infty.$$

This implies that the map $Z(h, w)$, defined on $H \times W$ by

$$(h, w) \to \sum_{k=0}^{\infty} \frac{1}{k!} (\nabla^k l(w), h^{\otimes k})_{H^{\otimes k}}$$

is almost surely well defined and $Z(h, w) = l(w + h)$ almost surely. $\square$

The following result is proven in [14]; for the sake of completeness, we give a short proof here:

Proposition 2. *Let $\hat{X}$ be defined as $\hat{X}(Y) = E[X|Y]$ where $E[X|Y]$ denotes the conditional expectation of X given Y. Then, for any $n \geq 1$, we have*

$$\nabla^n l = \widehat{X^{\otimes n}} l,$$

almost surely, where $\widehat{X^{\otimes n}}$ is defined as $\widehat{X^{\otimes n}}(Y) = E[X^{\otimes n}|Y]$ and $l(w)$ is defined as the expectation of $L(x, w)$ w.r.t. the measure $\alpha(\mathrm{d}x)$.

Proof. Let $\xi \in \mathbb{D}(H)$ be any test function; it follows then from the Cameron–Martin formula for μ that

$$\begin{aligned}
E\left[(\nabla l(Y), \xi(Y))_H\right] &= E\left[l(\nabla l, \xi)_H\right] \\
&= E\left[l(w) \int (x, \xi(w))_H L(x, w) \alpha(\mathrm{d}x)\right] \\
&= E\left[l(X+w)(X, \xi(X+w))_H\right] \\
&= E\left[l(Y)(X, \xi(Y))_H\right] \\
&= E\left[l(Y)(E[X|Y], \xi(Y))_H\right],
\end{aligned}$$

where

$$L(x, w) = \exp\left\{(x, w)_H - \frac{1}{2}|x|_H^2\right\}.$$

For the higher order derivatives, the proof is similar. □

Corollary 1. *The density l is an H-convex function.*

Proof. It suffices to show that the second derivative of l is a positive operator-valued distribution. This follows easily from the fact that

$$\begin{aligned}
(\nabla^2 l, h \otimes h)_2 &= l(\widehat{X^{\otimes 2}}, h \otimes h)_2 \\
&= l E[(X, h)_H^2 | Y] \geq 0
\end{aligned}$$

almost surely for any $h \in H$, where $(\cdot, \cdot)_2$ denotes the Hilbert–Schmidt scalar product. □

It follows from Proposition 2 that, μ-almost surely, $\nabla \hat{X}$ is a symmetric positive operator. In fact we have

Theorem 4. *There exists an H-convex and H-analytic function*

$$F \in \mathbb{D} = \bigcap_{p>1} \bigcap_{k \in \mathbb{N}} \mathbb{D}_{p,k},$$

such that $\hat{X}(w) = \nabla F(w)$ μ-almost surely.

Proof. It follows from Proposition 2 that $\nabla l = \hat{X} l$; taking the Sobolev derivative of both sides we get

$$\begin{aligned}
\nabla^2 l &= \nabla \hat{X} l + \hat{X} \otimes \nabla l \\
&= \nabla \hat{X} l + \hat{X} \otimes \hat{X} l. \qquad (3.3)
\end{aligned}$$

Again from Proposition 2, we have

$$\nabla^2 l = \widehat{X^{\otimes 2}} l, \qquad (3.4)$$

and since $l > 0$ μ-almost surely, we obtain from the relations (3.3) and (3.4)

$$\nabla \hat{X} = \widehat{X \otimes X} - \hat{X} \otimes \hat{X}, \tag{3.5}$$

and hence $\nabla \hat{X}$ is in the intersection of the spaces $\{L^p(\mu, H \otimes H),\ p > 1\}$; iterating this procedure we see that, for any $k \in \mathbb{N}$, $\nabla^k \hat{X}$ is in all L^p spaces, and hence it is in $\mathbb{D}(H)$. Moreover, from the Jensen inequality and from the equality (3.5), we see that

$$(\nabla \hat{X} h, h)_H = E[(X, h)_H^2 | Y] - E[(X, h)_H | Y]^2 \geq 0$$

for any $h \in H$ almost surely. Hence it is, μ-almost surely, a positive, symmetric operator on H. The Poincaré decomposition of vector fields on (W, H, μ) (which is also valid for the elements of $\mathbb{D}'(H)$; cf. [11]) says that

$$\hat{X} = \nabla F + \alpha,$$

where $F \in \mathbb{D}$, $\alpha \in \mathbb{D}(H)$ with $\delta\alpha = 0$. Hence

$$\nabla \hat{X} = \nabla^2 F + \nabla \alpha.$$

Since, from the equality (3.5), $\nabla \hat{X}$ is symmetric and since $\nabla^2 F$ is also symmetric, it follows that $\nabla\alpha$ is a symmetric operator; hence $\nabla\alpha = \nabla\alpha^\star$, where $\nabla\alpha^\star$ denotes the adjoint of $\nabla\alpha$ in H. Therefore

$$\begin{aligned} E[(\delta\alpha)^2] &= E[|\alpha|_H^2] + E[\mathrm{trace}\,(\nabla\alpha \cdot \nabla\alpha)] \\ &= E[|\alpha|_H^2] + E[\mathrm{trace}\,(\nabla\alpha^\star \cdot \nabla\alpha)] \\ &= 0. \end{aligned}$$

Since all the terms are positive, we should have $\nabla\alpha = 0$; hence α is a constant, and hence it is null. Consequently $\hat{X} = \nabla F$; moreover $\nabla^2 F$ is a positive operator, and therefore F is an H-convex function, i.e., $h \to F(w + h)$ is convex on H almost surely. □

Remark 4. We have also the following important information which is a consequence of Theorem 4:

1. It follows from $\nabla l = l\,\nabla F$ that $\mathcal{L}F = \mathcal{L}\log l$; hence $\log l = F - c$ and

   $$l = \exp\{F - c\},$$

   where $c = \log E[\mathrm{e}^F]$.
2. Note that, if we replace in the model $Y = X + w$ with $Y = \lambda X + w$, where $\lambda \in \mathbb{R}$ is a parameter, the same reasoning implies that the function λF is convex.
3. It follows from [3] that F has a modification with respect μ, which is a convex function on W in the ordinary sense.

Here are some further remarks:

Proposition 3. *We have the following identities:*

$$\begin{aligned} E[|X|_H^2] - E[|\hat{X}|_H^2] &= \int \mathrm{d}_H^2(\mu, \mu_{Y|X})\mu_X(\mathrm{d}x) - \mathrm{d}_H^2(\mu, (I_W + \hat{X})\mu) \\ &= 2\int L(x, w)\log L(x, w)\mathrm{d}\mu\mathrm{d}\alpha - \mathrm{d}_H^2(\mu, (I_W + \hat{X})\mu), \end{aligned}$$

and

$$E[|X|_H^2] - E[|\hat{X}(Y)|_H^2] = \int \mathrm{d}_H^2(\mu, \mu_{Y|X})\mu_X(\mathrm{d}x) - \mathrm{d}_H^2(l\mu, (I_W + \nabla F)l\mu).$$

Proof. Since $\hat{X} = \nabla F$, it is clear from [5] that the measure on $W \times W$ defined by $(I_W \times (I_W + \hat{X}))\mu$ is the unique solution of the MKP on the set $\Sigma(\mu, (I_W + \hat{X})\mu)$. Hence we have

$$E[|\hat{X}|_H^2] = \mathrm{d}_H^2(\mu, (I_W + \hat{X})\mu).$$

Moreover, we have (cf. [5])

$$\begin{aligned}\mathrm{d}_H^2(\mu, \mu_{Y|X}) &= |X|_H^2 \\ &= 2\int_W L(X, w)\log L(X, w)\mu(\mathrm{d}w),\end{aligned}$$

where the last equality follows from the Cameron–Martin theorem. Consequently

$$E[|X|_H^2] = \int \mathrm{d}_H^2(\mu, \mu_{Y|X})\mathrm{d}\mu_X. \quad \square$$

Proposition 4. *We have*

$$\mathrm{d}_H^2(\mu, \mu_Y) \leq \int_H |x|_H^2 \alpha(\mathrm{d}x) < \infty,$$

and consequently, by Theorem 3, there exists a 1-convex function $A \in \mathbb{D}_{2,1}$ such that $\mu_Y = l \cdot \mu = (I_W + \nabla A)\mu$ and the measure defined on $W \times W$ as $(I_W \times (I_W + \nabla A))\mu$ is the unique solution of the MKP(μ, μ_Y).

Proof. Recall that $\mu_Y = l \cdot \mu$ (i.e., $\mathrm{d}\mu_Y = l\mathrm{d}\mu$). It follows from the inequality of Theorem 1, Jensen's inequality and the Cameron–Martin theorem that

$$\begin{aligned}\mathrm{d}_H^2(\mu, l \cdot \mu) &\leq 2E[l \log l] \\ &\leq 2\int_{H\times W}\left[(w, x)_H - \frac{|x|_H^2}{2}\right]L(x, w)\alpha(\mathrm{d}x)\mu(\mathrm{d}w) \\ &= 2\int_{H\times W}\left[(w + x, x)_H - \frac{|x|_H^2}{2}\right]\alpha(\mathrm{d}x)\mu(\mathrm{d}w) \\ &= \int_H |x|_H^2\alpha(\mathrm{d}x) < \infty.\end{aligned}$$

Consequently μ and μ_Y are at finite Wasserstein distance from each other and the proof is then completed by an application of Theorem 3. $\square$

Let us calculate also the image of the measure $l \cdot \mu$ under the map $T = I_W + \nabla F$:

Theorem 5. *There exists some $G \in \mathbb{D}_{2,2}$, which is* 1*-convex and H-concave such that, for any $f \in C_b(W)$, one has*

$$\int_W f \circ T l\mathrm{d}\mu = \int_W f(w)\, l \circ T^{-1}(w)\frac{\mathrm{d}T\mu}{\mathrm{d}\mu}(w)\mathrm{d}\mu(w),$$

where

$$\frac{\mathrm{d}T\mu}{\mathrm{d}\mu} = \Lambda(G) = \det_2(I_H + \nabla^2 G)\exp\left(-\mathcal{L}G - \frac{1}{2}|\nabla G|_H^2\right).$$

Besides, G satisfies the following relations:

$$F(w) + G(y) + \frac{1}{2}|w-y|_H^2 \geq 0$$

for any $(x, y) \in W \times W$ *and*

$$F(w) + G(y) + \frac{1}{2}|w-y|_H^2 = 0$$

γ*-almost surely, where* γ *is the measure* $(I_W \times T)\mu$*. Consequently*

$$\frac{\mathrm{d}T(l.\mu)}{\mathrm{d}(l.\mu)} = \Lambda(G)\frac{l \circ (I_W + \nabla G)}{l}.$$

Proof. Since F is H-analytic and H-convex, we know already that T is invertible and its inverse is of the form $I_W + \nabla G$, where $G \in \mathbb{D}_{2,1}$ is the dual potential function of the MKP for $\Sigma(\mu, T\mu)$. Hence G is a 1-convex function; besides, the facts that F is H-analytic and that $h \to h + \nabla^2 F(w)h$ is invertible for almost all $w \in W$, where the negligible set is independent of the choice of $h \in H$, imply that G is a CH^∞-map. The H-convexity of F implies that ∇^F is a positive operator; hence

$$\nabla^2 G(y) = (I_H + \nabla^2 F)^{-1} \circ T^{-1}(y) - I_H$$

is a negative operator. Consequently G is H-concave (cf. [3]); this implies in particular that ∇G is a 1-Lipschitz map in the H-direction (cf. [13]). The rest of the proof follows from the change of variables formula on the Wiener space (cf. [13], Chapter 3). □

4. Estimations of divergence

Assume that the signal X is defined on another Wiener space on which the Sobolev derivative and the divergence operator are denoted respectively by $\tilde{\nabla}$ and $\tilde{\delta}$. In this setting it is useful to have an expression for $E[\tilde{\delta}m|Y]$, for $m \in \tilde{\mathbb{D}}_{p,1}(H)$, where the latter denotes the Sobolev space on this new Wiener space.

Proposition 5. *Let m be as above; then we have the following identity:*

$$E[\tilde{\delta}m|Y] = E[(\tilde{\nabla}_m X, Y - X)_H|Y],$$

almost surely.

Proof. Let $a \in \mathbb{D}$ and denote by θ the generic point of the Wiener space on which X is defined. Using the partial integration by parts formula, we have

$$\begin{aligned} E[\tilde{\delta}ma(Y)] &= E[\tilde{\delta}m(\theta)a(X(\theta)+w)] \\ &= E[\tilde{\delta}m(\theta)L(X(\theta,w))a(w)] \\ &= E[(m(\theta), \tilde{\nabla}L(X(\theta,w)))_H a(w)] \\ &= E\left[\left(m(\theta), (\tilde{\nabla}X(\theta), w)_H - \frac{1}{2}\tilde{\nabla}|X(\theta)|_H^2\right)_H L(X(\theta), w)a(w)\right] \end{aligned}$$

$$
\begin{aligned}
&= E[((\tilde{\nabla}_m X(\theta), w)_H - (\tilde{\nabla}_m X(\theta), X(\theta))_H) L(X(\theta), w) a(w)] \\
&= E[((\tilde{\nabla}_m X(\theta), w + X(\theta))_H - (\tilde{\nabla}_m X(\theta), X(\theta))_H) a(w + X(\theta))] \\
&= E[((\tilde{\nabla}_m X, Y)_H - (\tilde{\nabla}_m X, X)_H) a(Y)] \\
&= E[(\tilde{\nabla}_m X, Y - X)_H a(Y)]
\end{aligned}
$$

and the proof follows. □

Remark 5. Note that in the last line of the above equalities, the term with the scalar product is defined in the probabilistic sense as the limit

$$(\tilde{\nabla}_m X, Y - X)_H = \lim_{n\to\infty} (\tilde{\nabla}_m X, \pi_n w)_H$$

which is independent of the choice of the sequence $(\pi_n, n \geq 1)$ due to the independence.

It is also interesting to calculate an estimate of the divergence of the vector fields which are defined with respect to the initial Wiener measure:

Proposition 6. *Let $\eta \in \mathbb{D}_{p,1}$ for some $p > 1$ and denote the estimate of η as $\hat{\eta}(Y) = E[\eta|Y]$. We then have the following relation:*

$$
\begin{aligned}
E[\delta\eta|Y] &= (\delta\hat{\eta})(Y) - (\nabla F(Y), \hat{\eta}(Y))_H \\
&= (\delta\hat{\eta})(Y) - (\hat{X}(Y), \hat{\eta}(Y))_H,
\end{aligned}
$$

almost surely.

Proof. Let $a \in \mathbb{D}$; since δ is the adjoint of ∇ with respect to μ, we have

$$
\begin{aligned}
E\left[\delta\eta(w)a(Y)\right] &= E\left[(\eta, \nabla a(Y))_H\right] \\
&= E\left[(E[\eta|Y], \nabla a(Y))_H\right] \\
&= E\left[\left(\hat{\eta}(Y), \nabla a(Y)\right)_H\right] \\
&= E\left[l\left(\hat{\eta}, \nabla a\right)_H\right] \\
&= E\left[a\delta(l\hat{\eta})\right] \\
&= E\left[al\left\{\delta\hat{\eta} - (\hat{\eta}, \nabla \log l)_H\right\}\right] \\
&= E\left[a(Y)\left\{\delta\hat{\eta}(Y) - (\hat{\eta}(Y), \nabla F(Y))_H\right\}\right],
\end{aligned}
$$

and since a is arbitrary, the proof is completed. □

Since the scalar product in Proposition 5 is a divergence with respect to μ, we have the following

Corollary 2. *Under the hypothesis and with the notation of Proposition 5, we have*

$$
\begin{aligned}
E[\tilde{\delta}m|Y] &= E[(\tilde{\nabla}_m X, Y - X)_H|Y] \\
&= E[\delta(\tilde{\nabla}_m X)|Y] \\
&= \delta(\widehat{\tilde{\nabla}_m X})(Y) - (\nabla F(Y), E[\tilde{\nabla}_m X|Y])_H,
\end{aligned}
$$

almost surely, where $\widehat{\tilde{\nabla}_m X}(Y)$ denotes $E[\tilde{\nabla}_m X|Y]$.

5. The Girsanov measures

Let T be the transformation $T(w) = w + \nabla F(w)$, where $\nabla F = \hat{X}$ μ- and μ_Y-almost surely. Since F is a smooth H-convex function (it can be chosen even H-analytic if μ_X has exponential moments), it follows immediately from [13] that T is an invertible and absolutely continuous map. Let $\Lambda = \Lambda_T$ be defined by

$$\Lambda = \det\nolimits_2(I_H + \nabla^2 F) \exp\left\{-\mathcal{L}F - \frac{1}{2}|\nabla F|_H^2\right\}.$$

We have

Proposition 7. *The paths $\{T(y), y \in W\}$ are Gaussian distributed under the measure*

$$\mathrm{d}\nu = \frac{\Lambda}{l}\mathrm{d}\mu_Y.$$

In other words the law of the path $y \to T(y)$ under the probability ν is equal to the Wiener measure μ. This means in particular that the law of the random variable $\omega \to Y(\omega)+\nabla F(Y(\omega))$ is Gaussian under the initial probability measure.

Proof. Note that $T = I_W + \nabla F$ is a 1-monotone, H–$C1$-shift on the Wiener space W. It follows from Chapter VI of [13] that $T\mu$ is equivalent to μ and T has a two-sided inverse S such that

$$\mu\left(\{w \in W : T \circ S(w) = S \circ T(w) = w\}\right) = 1.$$

Since μ_Y is equivalent to μ, the same relation holds also when we replace μ with μ_Y. Let $f \in C_b(W)$; then, it follows again from Chapter VI of [13] that

$$\begin{aligned} E_Y\left[f \circ T \frac{\Lambda}{l}\right] &= E[f \circ T \Lambda] \\ &= E[f], \end{aligned}$$

where E_Y denotes the expectation with respect to the measure μ_Y. □

Remark 6. Note that the function F is also the unique convex (up to a constant) solution of the Monge–Ampère equation with transformation $T = I_W + \nabla F$:

$$\Lambda(M \circ T) = 1$$

μ-almost surely, where

$$M = \frac{\mathrm{d}T\mu}{\mathrm{d}\mu}.$$

Without loss of generality, we may assume that we are working on the classical Wiener space (cf. [13], to see how to insert the notion of time on an abstract Wiener space), i.e., $W = C_0([0, 1], \mathbb{R})$. Let us denote by $(\mathcal{F}_t^T, t \in [0, 1])$ the filtration generated by T, i.e., $\mathcal{F}_t^T = \sigma\{T_\tau(Y), \tau \le t\}$. Here it is important to note that Y represents the observation process. We have, using the Itô–Clark representation theorem for the Brownian motion $(t, y) \to T_t(y)$, where y represents the paths of Y,

$$\begin{aligned} f \circ T(Y) &= E_\nu[f \circ T(Y)] + \int_0^1 E_\nu[(D_t f) \circ T(Y) | \mathcal{F}_t^T] \mathrm{d}T_t(Y) \\ &= E[f] + \int_0^1 E_\nu[(D_t f) \circ T(Y) | \mathcal{F}_t^T] \mathrm{d}T_t(Y), \end{aligned}$$

for any $f \in \mathbb{D}_{p,1}$, $p > 1$.

References

[1] D. Feyel, A survey on the Monge transport problem, 2004. Preprint.

[2] D. Feyel, A. de La Pradelle, Capacités gaussiennes, Annales de l'Institut Fourier 41 (f.1) (1991) 49–76.

[3] D. Feyel, A.S. Üstünel, The notion of convexity and concavity on Wiener space, Journal of Functional Analysis 176 (2000) 400–428.

[4] D. Feyel, A.S. Üstünel, Transport of measures on Wiener space and the Girsanov theorem, Comptes Rendus Mathématiques 334 (1) (2002) 1025–1028.

[5] D. Feyel, A.S. Üstünel, Monge–Kantorovitch measure transportation and Monge–Ampère equation on Wiener space, Probability Theory and Related Fields 128 (2004) 347–385.

[6] D. Feyel, A.S. Üstünel, Monge–Kantorovitch measure transportation Monge–Ampère equation and the Itô calculus, in: Advanced Studies in Pure Mathematics, vol. 41, Mathematical Society of Japan, pp. 49–74.

[7] D. Feyel, A.S. Üstünel, The strong solution of the Monge–Ampère equation on the Wiener space for log-concave densities, Comptes Rendus Mathématiques, Ser. I 339 (1) (2004) 49–53.

[8] P. Malliavin, Stochastic Analysis, Springer-Verlag, 1997.

[9] T. Rockafellar, Convex Analysis, Princeton University Press, Princeton, 1972.

[10] M. Talagrand, Transportation cost for Gaussian and other product measures, Geometric and Functional Analysis 6 (1996) 587–600.

[11] A.S. Üstünel, Introduction to Analysis on Wiener Space, in: Lecture Notes in Math., vol. 1610, Springer, 1995.

[12] A.S. Üstünel, Analysis on Wiener space and applications. Electronic text at the site: http://www.finance-research.net/.

[13] A.S. Üstünel, M. Zakai, Transformation of Measure on Wiener Space, in: Springer Monographs in Mathematics, Springer-Verlag, 1999.

[14] M. Zakai, On mutual information, likelihood-ratios and estimation error for the additive Gaussian channel, IEEE Transactions on Information Theory 51 (9) (2005) 3017–3024.